Comparisons between duplicated genes have shown that gene conversions play an important role in the evolution of multigene families. Previous comparisons have documented in the recently duplicated γ-fetal globin genes of catarrhine primates, over 15 separate conversions affecting extensive stretches of coding and noncoding sequences. In the present study, δ- and β-globin genes from a lower primate Tarsius syrichta, and the δ-globin gene of the Asian great ape, Pongo pygmaeus, have been isolated and sequenced. Comparisons of these sequences with other primate δ and β sequences confirmed a previously reported conversion in an anthropoid ancestor and revealed additional conversions in basal primate, stem haplorhine, tarsier, and early lemur lineages. Conversions found between primate δ- and β-globin genes contrast with those found in the γ-genes in that δ-β conversions appear much less frequently and are more restricted to regions conserved by selection (i.e. coding and 5'-regulatory sequences). These differences indicate that soon after a duplication occurs, conversions can be quite frequent and encompass extensive portions of the duplicated region. With time, sequence differences accumulate, particularly in noncoding regions, and limit both the frequency and size of the conversions. Sequences conserved by selection accumulate differences more slowly and are therefore subject to gene conversions for a longer period of time. Both unconverted and converted sequences were consistent in supporting the placement of tarsier with anthropoids.

Studies of primate, lagomorph, rodent, artiodactyl, and marsupial β-globin genes (Hill et al., 1984; Jeffreys et al., 1982; Hardison, 1984; Hardies et al., 1984; Harris et al., 1983; Goodman et al., 1984; Koop and Goodman, 1988; Goodman et al., 1987) indicate that the ancestral β-globin cluster of eutherian mammals (65–85 million years ago) consisted of five linked genes (5'-γ-η-δ-β-3'). In the eutherian stem, ε, γ, and η-genes arose from an embryonically expressed progenitor while the δ- and β-genes arose from a postnatally expressed progenitor (Koop and Goodman, 1988). In mammals, the β-gene codes for the major adult β-chain and the δ-gene is either not expressed or expressed at reduced levels. Previous studies have shown that the δ-locus has not evolved as an independent lineage but in concert with the major adult β-globin genes. In each of the mammalian orders examined thus far, the δ-locus has acquired characteristics of the β-locus through a nonreciprocal exchange of nucleotide sequences or nonallelic gene conversion mechanism (Hardies et al., 1984; Hardison and Margot, 1984; Martin et al., 1983). These gene conversions have centered over the coding regions where they have essentially removed evidence for the early mammalian origins of the δ-locus. Evidence of early origins remains only in flanking and intron 2 sequences. In lemur, an unequal cross-over resulted in a hybrid γ-δ-locus (Jeffreys et al., 1982), but in most mammals the δ-locus evolved in close association with the adult β-locus.

Previous human and Old World monkey sequence comparisons used molecular clock models to suggest that a conversion of δ by β occurred in an anthropoid ancestor approximately 40 million years ago (Jeffreys et al., 1982; Martin et al., 1983; Hardison and Margot, 1984). In the present study we have extended these comparisons to include the complete sequences of an orangutan δ-gene and the δ- and β-genes of tarsier plus the recently sequenced lemur β (Harris et al., 1986) and spider monkey δ sequences (Spritz and Giebel, 1988). Comparison of these new sequences with orthologous sequences from humans and Old World monkeys confirms the previously reported conversion between the δ- and β-loci in an ancestral anthropoid approximately 40 million years ago (ancestral- or stem-anthropoid refers to the last common ancestor of hominoids, Old World monkeys, and New World monkeys). Our data further indicate that a more recent conversion between the δ- and β-loci occurred within the Tarsius lineage as late as 30 million years ago and that additional conversions occurred in very early primate δ and β evolution, most notably in ancestral haplorhines (anthropoid and tarsier common ancestor) and ancestral primates (haplorhine and lemuroid common ancestor). As in other mammalian lineages, these δ-β conversions are concentrated over the conserved coding and 5'-regulatory regions. Sequence relationships as determined from parsimony analyses of converted and nonconverted sequences clearly support the Haplorhini (anthropoids and Tarsius)/Strepsirhini (lorisoids and lemuroids) taxonomic division within Primates (Hill, 1955).
Tarsius $\delta$- and $\beta$-Globin Genes: Conversion and Evolution

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases: Accl, AvalI, BamHI, BglII, BstEII, EcoRl, HindIII, MboI, NcoI, Stul, PstI, PvuII, and Xbal were obtained from New England Biolabs. T4 ligase, polynucleotide kinase, and bovine alkaline phosphatase were obtained from New England Biolabs or United States Biochemical Corp. Radioactive nucleotides were obtained from ICN. Chemicals for DNA sequencing were obtained from vendors recommended by Maxam and Gilbert (1980). X-ray film (XAR) was obtained from Kodak. Nylon filters for blots were obtained from Schleicher and Schuell and No. 3MM absorbent paper from Whatman.

DNA Isolation, Cloning, and Sequencing—High molecular weight DNA was prepared from Tarsius syrichta liver tissue (Blin and Stafford, 1976) obtained from the tarsier colony at the Duke University Primate Center. Genomic DNA libraries were constructed in Charon 35 (Loenen and Blattner, 1983) and 40 (Dunn and Blattner, 1987) bacteriophage vectors, propagated in K-12 strain EDS767 rec A- host (Murray et al., 1977), and screened for $\beta$-globin genes by hybridization with labeled human $\epsilon$, $\gamma$, and $\beta$-globin genes (Benton and Davis, 1977). Recombinant $\lambda$ clones containing $\beta$-globin gene sequences were isolated and mapped with restriction enzymes BamHI, EcoRI, and HindIII. Genes were identified by comparison of cloned and genomic DNA restriction fragments that hybridized to human $\epsilon$, $\gamma$, and $\beta$-gene probes (Southern, 1975). Genotype-containing fragments were subcloned into pUC19 plasmid vectors, transformed, and replicated in Escherichia coli K-12 strain JM109 hosts. Plasmid subclones (from Ch32 14.2) containing the orangutan $\beta$-gene (Koop et al., 1986b; Slightom et al., 1987) were also examined. Nucleotide sequences were obtained using the method of Maxam and Gilbert (1980) as modified by Slightom et al. (1987, 1988). Over 90% of transcribed sequences were sequenced on both strands, and all of the sequences were sequenced on at least two independently labeled DNA fragments.

Comparative Analysis—DNA sequences were aligned with other $\delta$ and $\beta$ sequences by comparing pairwise alignments and combining the results by hand. The results of several alignment algorithms (such as that of Wilbur and Lipman, 1983; Lipman and Pearson, 1985; Zweig, 1984) were obtained using the method of Smith and Waterman, 1981; Wilbur and Lipman, 1983; Lipman and Pearson, 1985) nor dot matrix comparisons (Zweig, 1984; 13 of 20 matches) permit the complete alignment of $\delta$ with $\beta$ sequences. Only exon, intron 1, and $5'$-regulatory sequences of $\delta$ and $\beta$ could be aligned with one another. Intron 2 and flanking sequences were clearly too distantly related to be compared.

For sequence divergence analyses, noncoding sequence divergence was calculated directly, counting all substitutions, insertions, and gaps as single events. The number of substitutions per site was determined by the principle of minimum evolution to be the best working hypothesis of genealogical relationships. The advantage of the parsimony method is that information at each variable position can be evaluated separately. The additional use of clearly defined outgroups further enables the direction of individual mutations to be estimated for each aligned position within the study group. For example, while the framework established by the parsimony criteria, a base shared between a reference species such as the rabbit or goat and one or more primate species is assumed to reflect a primitive condition. Any changes from that primitive base is therefore a derived condition. Genealogical relationships are based only on shared derived features. Another feature of parsimony analyses is that ancestral sequences are estimated as part of the parsimony procedure. Thus if all primate $\beta$-genes shared one base and all nonprimate $\delta$ and all mammalian $\delta$-genes shared another base, parsimony would place a single mutation in the ancestral primate $\beta$-lineage that could be used to support single origins of all primate $\beta$-genes.

Identification of Converted Regions—One of the most successful ways of identifying gene conversions in mammalian multigene families is by noting differences in geography for a specific region are...

---

**Fig. 1.** Tarsius $\delta$-globin cluster organization and structure compared with that of the brown lemur and owl monkey. The dashed lines in the $\beta$-gene map of Tarsius indicate regions where linkage is suggested only from the hybridization of specific probes to genomic blots. The square brackets in the lemur $\beta$-gene map indicate where a deletion in the lemur linkage occurred. Restriction endonuclease maps of the six recombinant $\lambda$ clone inserts are positioned with respect to the map of the Tarsius $\delta$-globin gene cluster. Restriction maps of the three plastidial clones used to sequence tarsier $\delta$- and $\beta$-genes and the sequencing strategy are presented below the $\lambda$ clones. Restriction endonucleases: $B$ = BamHI, $E$ = EcoRI, $H$ = HindIII, $N$ = NcoI, $P$ = PstI, $S$ = Stul, $X$ = Xbal.
inconsistent with phylogenies for adjacent sequences and well established species phylogenies. This requires an understanding of which species are more closely related than others. As there is only one true species phylogeny, the evolution of all orthologous sequences (sequences in different species descended from the same, most recent duplication product; Fitch, 1977) should reflect the same species phylogeny. If a particular sequence phylogeny does not reflect the same species phylogeny, then the existence of such events becomes a reasonable hypothesis. For example, previous comparisons of flanking and intron 2 sequences of and and -genes from primates, lagomorphs, and rodents and flanking sequences from artiodactyls indicate that the and -loci arose from a duplication occurring prior to the separation of these four mammalian orders, but regulatory, coding, and intron 1 sequences suggest that the -locus was the product of recent duplications occurring within each of the four mammalian orders. In this case the hypothesis of a gene conversion event in each of the four mammalian orders has the effect of reconciling two apparently incompatible sequence phylogenies and providing a valuable insight into the role of gene conversion in the evolution of multigene families (Jeffreys et al., 1982; Martin et al., 1983; Hardison and Margot, 1984; Hardison, 1984; Hill et al., 1984; Hardies et al., 1984). Using this same logic in comparing the two -genes of human, chimpanzee, gorilla, orangutan, and rhesus monkey, over 15 different gene conversions have been postulated to have occurred between the duplicated -genes of these catarrhine primates (Slightom et al., 1985, 1987, 1988) and in a study of six different human -gene alleles as many as 13 conversion events may have occurred just within the human lineage (Powers and Smithies, 1986). A more detailed description of the parsimony procedure and the use of parsimony analyses in locating sequences that have undergone gene conversions is described elsewhere (Goodman et al., 1979; Wiley, 1981; Slighom et al., 1987, 1988).

RESULTS

Tarsier -Globin Cluster Structure and Organization—Two genomic libraries from T. syrichta liver tissue (Charon 35 vector, 1.5 x 10⁶; and Charon 40 vector, 5 x 10⁶ titers) were screened with human -, - and -gene probes, and six recombinant clones were isolated. These clones were mapped by partial sequencing of gene regions. Linkage between the and -loci, though not confirmed, is suggested by sizes of

---

**Fig. 2.** Aligned 6-nucleotide sequences. Human (Hsa, Homo sapiens), orangutan (Ppy, Pongo pygmaeus), rhesus macaque (Mmu, Macaca mulatta), colobus (Cpo, Colobus polykomos), baboon (Pan, Papio anubis), and spider monkey (Ateles geoffroyi), tarsier (Tey, T. syrichta), lemur (Lfu, Lemur fulvus), and Australian rabbit (Ocu, Oryctolagus cuniculus). The three-letter designation is an abbreviated form of the binomial species name (e.g. Hsa, Homo sapiens). References for each of these sequences are given in the text. Gaps (**) were introduced to improve alignments. Only variable positions are given below the complete human - sequence. The sequence of a tarsier and and are shown below the alignment at position 1612 by an "alu." Enzyme and sequence patterns typical of promoter sequences (Myers et al., 1986; Maniatis et al., 1987) are noted above the human - sequence. The nucleotide sequence of the orangutan -gene indicates that the orangutan -chain differs from the human -chain (which codes for the -chains in the minor (A2) hemoglobin component) by one amino acid at position 126 of the amino acid chain (Val in orangutan and Met in humans).
Tarsius and β-Globin Genes: Conversion and Evolution

Fig. 2—continued.
overlapping HindIII, EcoRI, and BamHI fragments detected in genomic and cloned DNAs. The general organization of the Tarsier β-globin cluster closely follows that of owl monkey (Harris et al., 1986) and the proposed primitive eutherian mammal β-cluster (Bunn and Forget, 1986; Goodman et al., 1984, 1987; Koop and Goodman, 1988). It does not include the hybrid v-β-gene found in the brown lemur (a prosimian representative).

**Tarsier δ- and β-Gene Sequences**—Tarsier δ- and β-genes were completely sequenced along with two previously isolated EcoRI fragments containing the orangutan β-gene. These sequences are aligned with other primate δ- and β-genes in Figs. 2 and 3. The predicted amino acid sequence of the T. syrichta β-chain agrees very closely with the known sequence of Tarsius bancanus (Beard et al., 1976, Beard and Goodman, 1976). The β-chain amino acid sequence of the two Tarsius species differs at positions 56, 73, and 141 where T. syrichta has Ser, Asp, and Phe, and T. bancanus has Gly, Gln, and Leu, respectively. The minor hemoglobin component of adult T. bancanus, representing 18% of the total (Barnicot and Hewett-Emmett, 1974), was shown to contain an altered β-chain that shared Ala, His, and His at positions 5, 116, and 117 with the major β-chain (Beard and Goodman, 1976). These amino acids also...

---

**Fig. 3.** Aligned β-nucleotide sequences. Human (Hsa), chimpanzee (Ptr, Pan troglodytes), gorilla (Ggo, Gorilla gorilla), macaque (Mcy, Macaca cynomolgus), colobus (Cpo), tarsier (Tsy), lemur (Lf), Australian rabbit (Ocu), European hare (Leu, Lepus europaeus), and goat (Chi, Capra hircus). Gaps were introduced to improve alignments and only variable positions are given below the complete human β sequence. An Alu sequence located at position 745 is given below the complete sequence. Exons and promoter sequences are noted above the human (3′ sequence. Sequences of tarsier δ- and β-genes contain promoter sequences of the type and location found in other 6- or β-genes (Myers et al., 1986; Maniatis et al., 1987; Martin et al., 1983), as well as functional coding and splicing sequences (Breathnach et al., 1978). The δ- and β-globin genes differ with respect to 5′-promoter sequences. Specifically, δ-genes lack the CACCC sequences found in β-genes and have three possible CCAAT sequences located 120, 140, and 205 base pairs upstream of the initiation site (Fig. 2). Only the most 3′ CCAAT variant appears to be homologously related to the CCAAT element of the β-gene. Other than the TATA sequence, conserved 5′ sequences revealing positions of promoter sequences in δ-genes are conspicuously absent in β-genes. Differences in 5′-promoter regions do not appear to be the only factor in explaining reduced expression levels of the δ-gene. Undefined sequences within intron 2 of the β-gene also appear to increase transcription rates (Kosche et al., 1985). The longest stretch of conserved sequence in intron 2 of β also appear to increase transcription rates (Kosche et al., 1985). This sequence (GTTTGAAT) does not resemble any known enhancer sequence and is not found in other globin genes. Its conservation among the β-globin genes of different eutherian orders suggests an important role, perhaps in regulation.
reside at corresponding positions of the deduced T. syrichta δ-chain. In addition, amino acid compositions of selected trypic peptides from major and minor chains of T. bancanus (Beard and Goodman, 1976) indicated identity between these chains at positions 1-8 and 112-120. The amino acid sequences of T. syrichta δ- and β-chains as deduced from nucleotide sequences (Figs. 2 and 3) indicate only a single difference (position 6) over these same positions. Electrophoresis of T. syrichta hemoglobin proteins confirms the presence of a slower migrating (at pH = 8.6) minor hemoglobin component like that of T. bancanus (data not shown). This slower hemoglobin is consistent with the reduced number of negatively charged amino acids found in the δ-chain (see below). It is therefore very likely that the δ-locus of T. syrichta is expressed as in T. bancanus and that the δ-locus codes for the altered β-chain present in the minor hemoglobin component. T. syrichta δ- and β-chains differ by only six amino acids (at positions 6, 16, 19, 121, 126, and 139, δ has Asp, Ser, Asn, Gln, Leu, and Ala and β has Glu, Gly, Asp, Ala, Glu, and Val, and Thr, Figs. 2 and 3).

Divergence Analysis—Neither pairwise alignment comparisons (Smith and Waterman, 1981; Wilbur and Lipman, 1983; Lipman and Pearson, 1985) nor dot matrix comparisons (13 matches out of 20 in a 4-fold compression, Zweig, 1984) revealed any evidence of sequence homology between the flanking and intron 2 sequences of δ- and β-loci. This together with the fact that, for these regions, alignments between primate and lagomorph δ and primate and lagomorph β sequences are easily obtained (Figs. 2 and 3) clearly demonstrates that the δ- and β-loci originated prior to the separation of primate and rabbit lineages. Only in the region between the CCAAT promoter and the end of exon 2 (positions 291-885 in Fig. 2 and positions 1208-1800 in Fig. 3) and in exon 3 were alignments between the two loci possible. As the ability to detect conversions occurs only where common alignments between δ and β sequences can be obtained, these regions were analyzed separately from those regions in which δ and β sequences could not be aligned and were thus clearly part of two anciently separated paralogous lineages.

Divergence among primate flanking and intron 2 sequences of δ- and β-genes (Table I) indicates that cercopithecoid (Old World monkeys) and hominoid sequences are most closely related (divergence is about 5-7%) and ceboid (New World monkeys) sequences diverge about 10-11% from catarrhine (hominoid and cercopithecoid) sequences. Lemur, tarsier, and anthropoid (catarrhine and platyrrhine) sequences are all about equally divergent from each other (divergence is about 24-28%). As inferred from sequence divergence, hominoids are most closely related to cercopithecoids, followed by ceboids and then tarsioids and lemuroids, with anthropoid, tarsioid, and lemuroid relationships being unresolved. These same relationships are evident in both δ- and β-gene lineages and are compatible with conclusions based on a sizable body of morphological, amino acid, and nucleotide sequence, and DNA hybridization data (Koop et al., 1986a; Bonner et al., 1980; Schwartz, 1986; Miyamoto and Goodman, 1986).
The sequence divergences among δ and β coding, 5'-regulatory, and intron 1 regions from different groups of primates, however, are not indicative of two anciently separated gene lineages (Table II). This is particularly true in the noncoding and synonymous coding sequences. Because δ- and β-genes are expressed differently in separate primate lineages, selective pressures affecting the number of amino acid changing substitutions may also be different. The δ-gene of catarrhines, for example, is completely silent whereas the δ-gene of tarsiers makes up about 18% of the total adult β-hemoglobin. The β-locus, however, is always the major adult hemoglobin. In noncoding and synonymous sequence divergence, anthropoid sequences are to tarsier or lemur sequences are more similar to each other than catarrhine (ceropithecoid and hominoid) sequences are from platyrrhine (New World monkeys) δ sequences. These patterns of divergence support the conclusions of Hardison (1984), Hardison and Margot (1984), and Martin et al. (1983) which contend that a conversion may have also occurred between ancient primate δ and β-genes (16.1%) was also less than that of catarrhines and platyrrhines (24.8%) and more than that between humans and baboons (10.5%); therefore a conversion in lemur may also have occurred about 30 million years ago. In 5'-regulatory and intron 1 sequences an additional pattern of δ and β sequence divergence appears. In these noncoding regions, all of the primate δ and β sequences are more similar to each other than primate δ sequences are to rabbit δ sequences or primate β sequences are to rabbit β sequences. This suggests that a conversion may have also occurred between ancient primate δ and β sequences prior to the separation of anthropoid, tarsiid, and lemuroid lineages (50–60 million years ago).

Table II
δ and β sequence divergence matrix

Nonsynonymous and synonymous δ and β coding sequence divergences (% uncorrected) are presented above the null diagonal (nonsynonymous divergence is given above synonymous divergence) and sequence divergences (% uncorrected) of δ'-regulatory and intron 1 regions are presented below the null diagonal. Pairwise sequence divergence values are obtained where the row number and the column number intersect.

<table>
<thead>
<tr>
<th>Exons 1, 2, and 3</th>
<th>Human β</th>
<th>-</th>
<th>3.2</th>
<th>4.5</th>
<th>10.6</th>
<th>5.1</th>
<th>3.3</th>
<th>5.1</th>
<th>3.2</th>
<th>5.7</th>
<th>7.3</th>
<th>18.3</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>10.2</td>
<td>32.2</td>
<td>28.1</td>
<td>27.0</td>
<td>19.7</td>
<td>24.5</td>
<td>21.7</td>
<td>26.4</td>
<td>23.6</td>
<td>36.7</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Macaque δ</td>
<td>3.5</td>
<td>-</td>
<td>6.8</td>
<td>10.4</td>
<td>7.2</td>
<td>3.6</td>
<td>6.1</td>
<td>4.1</td>
<td>7.5</td>
<td>10.2</td>
<td>16.8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>31.3</td>
<td>27.4</td>
<td>25.0</td>
<td>19.4</td>
<td>22.9</td>
<td>17.5</td>
<td>19.6</td>
<td>23.2</td>
<td>25.0</td>
<td>33.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tarsier δ</td>
<td>17.7</td>
<td>-</td>
<td>11.0</td>
<td>6.3</td>
<td>6.3</td>
<td>7.3</td>
<td>4.7</td>
<td>1.8</td>
<td>6.5</td>
<td>17.1</td>
<td>3.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>30.1</td>
<td>33.3</td>
<td>28.8</td>
<td>31.3</td>
<td>30.6</td>
<td>18.3</td>
<td>26.5</td>
<td>32.3</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemur δ</td>
<td>15.1</td>
<td>16.3</td>
<td>17.1</td>
<td>-</td>
<td>10.8</td>
<td>11.6</td>
<td>12.3</td>
<td>11.4</td>
<td>12.8</td>
<td>8.9</td>
<td>20.4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>23.8</td>
<td>29.8</td>
<td>29.7</td>
<td>25.8</td>
<td>21.2</td>
<td>16.1</td>
<td>33.1</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit β</td>
<td>27.6</td>
<td>27.6</td>
<td>30.0</td>
<td>25.3</td>
<td>-</td>
<td>7.2</td>
<td>8.4</td>
<td>6.5</td>
<td>7.5</td>
<td>5.9</td>
<td>14.9</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>29.2</td>
<td>29.3</td>
<td>29.2</td>
<td>30.3</td>
<td>30.3</td>
<td>26.0</td>
<td>33.8</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human δ</td>
<td>12.6</td>
<td>13.0</td>
<td>20.2</td>
<td>17.9</td>
<td>26.9</td>
<td>-</td>
<td>3.3</td>
<td>3.2</td>
<td>7.0</td>
<td>9.8</td>
<td>18.1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>10.5</td>
<td>20.1</td>
<td>23.4</td>
<td>22.9</td>
<td>31.4</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baboon δ</td>
<td>13.3</td>
<td>14.2</td>
<td>21.3</td>
<td>19.0</td>
<td>28.0</td>
<td>7.1</td>
<td>-</td>
<td>5.0</td>
<td>8.5</td>
<td>9.8</td>
<td>19.7</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>24.8</td>
<td>25.4</td>
<td>17.6</td>
<td>32.5</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spider M δ</td>
<td>16.5</td>
<td>16.9</td>
<td>22.1</td>
<td>21.5</td>
<td>30.4</td>
<td>12.8</td>
<td>13.1</td>
<td>-</td>
<td>5.9</td>
<td>7.9</td>
<td>17.9</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>23.9</td>
<td>25.6</td>
<td>36.0</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarsier δ</td>
<td>20.5</td>
<td>20.5</td>
<td>7.8</td>
<td>18.0</td>
<td>28.2</td>
<td>20.2</td>
<td>21.8</td>
<td>23.2</td>
<td>-</td>
<td>8.6</td>
<td>17.9</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>15.1</td>
<td>32.0</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemur δ*</td>
<td></td>
<td>-</td>
<td>17.7</td>
<td>32.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

* Divergence values based on partial sequence.
Parsimony Analysis—To further investigate conversions indicated by sequence divergences, we have examined each variable position of aligned $\delta$ and $\beta$ sequences and determined, on the basis of parsimony, which positions supported conversions and which positions did not. At each variable position where positional paralogues shared a sequence identity that parsimony depicted as derived rather than primitive, the result was taken as evidence for a conversion (for definitions and a description of the methods see "Experimental Procedures" and Slightom et al., 1987, 1988). Where positional orthologues shared a derived nucleotide sequence, the result was taken as evidence against conversion in those taxa sharing the derived change. For example, when primate $\delta$ and $\beta$ sequences are closer to each other than primate $\delta$ sequences are to rabbit $\delta$ sequences and primate $\beta$ sequences are to rabbit and goat $\beta$ sequences, a conversion event is possible. The results of all of the separate parsimonious solutions are presented in Fig. 4. A square box at a variable position indicates that positional paralogues are closer genealogically to each other than to positional orthologues, therefore support-

![Diagram](image-url)

**Fig. 4. Parsimony analysis of aligned $\delta$- and $\beta$-globin genes and the identification of gene conversion regions.** The aligned $\delta$- and $\beta$-genes of human, Old World monkey (Cercopithecus, includes colobus, macaques, and baboons), New World monkey (Ceboid, includes the spider monkey), tarsier, and lemur plus catarrhine (Catar), anthropoid (Anthr), haplorhine (Haplo), and primate ancestors were used to locate gene regions that may have been involved in conversion events. Above the organizational map of $\delta$- and $\beta$-globin genes (exons 1, 2, and 3) are indicated by raised bars) are the results of a position by position parsimony analysis of regions where $\delta$ and $\beta$ sequences could be aligned. Ancestral genes reconstructed by the parsimony procedure were also included in the analysis to determine when conversions occurred. Triangles indicate variable nucleotide sequence positions (positioned with respect to the gene map of $\delta$ and $\beta$) in which paralogous (separate $\delta$- and $\beta$-gene) lineages do not support a conversion hypothesis. Boxes indicate variable positions in which $\beta$-genes show a closer affinity to $\delta$-genes than to other $\beta$-genes (and vice versa) and therefore suggest a possible conversion event. Gene conversions are hypothesized where two or more adjacent squares are found. The lemur $\delta$, because it has been fused with the $\eta$-gene (indicated by the dashed arrow), retains only a small part of $\delta$ exon 2. A single position in exon 2 plus two positions in exon 3 supports a conversion event in lemur; therefore, a conversion event between lemur $\delta$ and $\beta$ sequences seems likely. Below the $\delta$- and $\beta$-gene maps is a regional analysis of $\delta$ and $\beta$ sequences ($H =$ human; $C =$ chimp; $G =$ gorilla; $M =$ macaque; $Co =$ colobus; $S =$ spider monkey; $O =$ orangutan; $P =$ baboon; $T =$ tarsier; $L =$ lemur; $R =$ European hare; $Gt =$ goat). Only in regions 2 and 4 could $\delta$ and $\beta$ sequences be aligned with one another. Regions 1, 3, and 5 of $\delta$- and $\beta$-genes could not be aligned and therefore cannot participate in possible conversions occurring within the primates (that primate and rabbit orthologous sequences could be aligned over regions 1, 3, and 5 supports this conclusion). The lack of two distinct $\delta$- and $\beta$-lineages dating prior to the separation of primates and rabbits in regions 2 and 4 contrast with the branching pattern shown in regions 1, 3, and 5. Gene conversion offers a reasonable explanation that could resolve the conflicting gene phylogenies and still be compatible with a single species phylogeny. Although a ceboid $\beta$-gene sequence is not known, the absence of nucleotide positions grouping spider monkey $\delta$ sequences with hominoid and catarrhine $\beta$ sequences indicates the absence of conversions between $\delta$ and $\beta$ in the ceboid lineage. However, the spider monkey $\delta$ sequence in region 2 does not clearly group with higher primate $\delta$ or $\beta$ sequences; therefore, either conversions may have occurred within the ceboid lineage or platyrrhine and catarrhine lineages separated soon after a conversion in their common anthropoid ancestor. Ceboid $\beta$ sequences are needed to ultimately resolve this question.
ing gene conversion. A triangle, on the other hand, indicates that positional orthologues are more closely related than parallelous sequences and hence do not support conversion. Stated in a more general way, triangles represent primate \( \beta \)-sequences that are more similar to other \( \beta \)-sequences than to \( \delta \)-sequences, and squares represent \( \delta \)-sequences that are more similar to other \( \delta \)-sequences than to other \( \beta \)-sequences. The same can be stated for \( \delta \)-sequences. On the basis of this analysis, conversions, hypothesized only where two or more boxes are clustered, have affected the region between the CCAAT element and the end of exon 2 (region 2) in the stem primates, stem haplorhines, stem anthropoids, lemur, and tarsier lineages. Part of exon 3 (region 4) also appears to be converted in stem haplorhine and in the lemur lineage, although evidence in this small region is not strong. That sequences 5’ of the CCAAT element and 3’ of exon 3 and intron 2 (regions 1, 3, and 5) were not affected by these conversions is evidenced by our inability to find an alignment between \( \delta \)- and \( \beta \)-genes in these regions.

Parsimony analysis of each of the five regions is presented in Fig. 4 below the gene map (phylogenetic patterns of regions 1, 3, and 5 were totally compatible and were therefore combined). Trees from regions 2 and 4, in contrast with trees for regions 1, 3, and 5 in that they do not show \( \delta \) and \( \beta \)-sequences evolving as two separate lineages originating prior to the separation of primates and lagomorphs. In regions 2 and/or 4, anthropoid \( \delta \)- and \( \beta \)-genes group together as do haplorhine, primate, tarsier, and lemur gene lineages. Phylogenetic trees for regions 2 and 4 are therefore compatible with conversions hypothesized from position analysis as well as those suggested by divergence analyses.

The branching arrangements indicated in Fig. 4 join hominoids to cercopithecoids followed by ceboids, tarsier, lemur, rabbit, and goat. As the phylogenetic position of tarsier has been the subject of considerable controversy (Poppock, 1918, Hill, 1955; Dene et al., 1976, Schwartz, 1986; Sarich and Cronin, 1976, Baba et al., 1982), we examined alternate placements of Tarsius. To place tarsiers with lemurs would require hypothesizing 12 additional events in regions 1, 3, and 5 and 2 additional events in region 4. Alternatively, to place Tarsius at a position ancestral to anthropoids and lemurs would require 21 additional events in regions 2 and 4 and in region 5 all alternatives are equally possible. Clearly these data support placing Tarsius \( \delta \) and \( \beta \)-sequences closer to orthologous anthropoid sequences than to prosimian sequences.

**DISCUSSION**

Until recently there has been no direct evidence for the expression of the \( \delta \)-locus in lower primates. Protein studies have shown only that two very similar \( \beta \)-like chains are expressed in tarsiers (\( \beta_{\text{maj}} = 82\% \), \( \beta_{\text{min}} = 18\% \); Beard et al., 1976) as well as galagos (\( \beta_1 = 60\% \), \( \beta_2 = 40\% \); Watanabe et al., 1985; Tagle et al., 1988). In this study we have demonstrated in Tarsius that the two \( \beta \)-chains are coded by a \( \beta \)-locus and a \( \delta \)-locus that has undergone gene conversion by the \( \beta \)-locus. There is no evidence for a duplicated \( \beta \)-gene. Southern blots of genomic galago DNA confirm the presence of five loci (Koop and Goodman, 1988; Tagle et al., 1988) of which three have been identified as embryonic type \( e \), \( \gamma \) and \( \tau \)-genes (Tagle et al., 1988), therefore the remaining two loci must code for the two adult \( \beta \)-hemoglobin chains. The very high amino acid sequence similarity between the two galago adult \( \beta \)-chains indicates either a recent duplication of the \( \delta \)-locus in addition to a deletion of the \( \delta \)-locus or a single conversion of the \( \delta \)-locus by \( \beta \). Given the history of the \( \delta \)-locus, we advocate the latter more parsimonious hypothesis.

We have also uncovered evidence for the involvement of conversions between the \( \delta \)- and \( \beta \)-loci prior to the formation of a hybrid \( \gamma \)-gene in lemurs. Both coding sequence divergence analysis (Table II) and parsimony analysis (Fig. 4, regions 2 and 4) suggest that a conversion affected at least part of exon 2 and exon 3 and \( \delta \). Harris et al. (1986) suggested that conversions between \( \delta \) and \( \beta \) occurring prior to the fusion of \( \gamma \)- and \( \delta \)-genes might help explain why the lemur \( \beta \)-chain appears to have accumulated such an inordinate number of amino acid substitutions. This hypothesis gains considerable support from the present data which indicates that such a conversion probably did occur.

Additional conversions of \( \delta \) by \( \beta \) in stem primates, stem haplorhines, and stem anthropoids are also supported by both sequence divergence (Table II) and parsimony analyses (Fig 4). These events are incorporated into a general scheme depicting some of the major events occurring in the evolution of the primate \( \beta \)-globin cluster (Fig. 5). Within primates as many as five gene conversions involving \( \delta \)- and \( \beta \)-genes have occurred.

As the \( \beta \)-locus codes for the primary adult \( \beta \)-type hemoglobin in all mammals, we and others have assumed that the evolutionary history of \( \delta \)- and \( \beta \)-genes have involved \( \delta \) being converted by \( \beta \). We cannot however dismiss the possibility that some changes originally introduced into the \( \delta \)-locus may be transferred to the \( \beta \)-gene as a result of these conversions. As suggested by Harris et al. (1986), sequence introgression from a less constrained \( \delta \)-locus into a \( \beta \)-gene may result in introducing increased \( \beta \)-variability. This type of mechanism might also explain a tendency noted by Spritz and Giebel (1988) for some synonymous divergence comparisons (Table II) to be more divergent than noncoding and flanking sequences (Table I). In general however it appears that that tighter selectional constraints acting on the \( \beta \)-locus determines, if not the direction of conversions themselves, at least which conversions survive and become fixed in populations.

Comparing the conversions between \( \delta \)- and \( \beta \)-genes with those between catarrhine \( \gamma \)-genes reveals several differences that may simply reflect the much older origins of the \( \delta \)- and \( \beta \)-genes (approximately 85–100 million years ago versus 25–35 million years ago; Slightom et al., 1980, 1985, 1987, 1988; Sheth et al., 1981). Conversions between \( \gamma \)-genes are much more frequent (as many as 13 conversions have occurred in the human lineage alone, Powers and Smithes (1986) and within other catarrhine species over 12, Slightom et al. (1988)) and are much more extensive (spanning extensive noncoding regions as well as coding sequences; Slightom et al., 1985, 1987, 1988; Scott et al., 1984). Conversions between primate \( \delta \)- and \( \beta \)-genes, as evident in this paper, are relatively frequent and are localized to regions maintaining, through selection, high levels of similarity (between the CCAAT promoter sequence and the end of exon 2 and exon 3). Conversions themselves contribute to higher levels of sequence matches but are not enough to maintain such similarity. That intron 2 lies between two areas of conversion but is not converted (Fig. 4) indicates an important role of selection in determining where conversions occur. It would appear then that soon after a gene duplication occurs, conversions can be frequent and extend over the entire duplicated region. Over time, differences accumulate, particularly in noncoding regions, and conversions become less frequent and limited to regions conserved by selection. These empirical observations are compatible with models constructed on a more theoretical basis (Walsh, 1987).

That conversions play a major role in the evolution of multigene families, such as the \( \beta \)-globin family has major
Repercussions when inferring species phylogeny from sequence phylogeny. Inconsistent phylogenetic patterns between different regions within a set of homologous sequences permit the detection of converted regions. Once these conversions are incorporated into the phylogenetic history of a sequence, there should be consistency among the different regions with respect to species relationships. This is in fact what we have found. All of the δ and β regions examined were compatible with hominoids being more closely related to cercopithecoids, followed by ceboids, tarsiers, lemurs, rabbits, and goats, respectively. This branching arrangement is quite strongly supported by both δ and β sequences and leads us to strongly favor the placement of Tarsius with anthropoids in Haplorhini rather than with lemurs and lorises in Prosimii.

Acknowledgments—We would like to thank Danilo Tagle and David Fitch for their help in isolating clones and for their comments on this manuscript.

REFERENCES
